

食品偽装防止に向けた取扱履歴の整合性判定による
トレーサビリティシステムの開発

山田詩穂[†] 石塚美伶[†] 原田翔貴[†] 竹原一駿[†] 鎌田典彦[‡] 喜田弘司[†]
香川大学[†] 日本電気通信システム株式会社[‡]

1. はじめに

近年、食品の産地・銘柄偽装により消費者の食品に対する信頼が揺らいでいる。この対策として生産から販売までの食品の取扱いの履歴(以下、取扱履歴)をブロックチェーンに記録し、提供するトレーサビリティシステムが注目されている。しかし、ブロックチェーンでは過去に登録された取扱履歴の改ざんは防止できるが、偽装された取扱履歴の登録は防止できない。そこで本稿では、産地・銘柄の組み合わせ毎の加工後の食品の重量を予測することで偽装を検出する方式を提案し、その方式を用いたトレーサビリティシステムを開発し実験を実施した。

2. 従来方式

松村らが提案するトレーサビリティシステムでは、食品の流通経路上の全ての事業者に入出荷時の食品の重量を登録させ、小売での流通量が生産地での生産量を越えていないかを確認することで、偽装を検出していた[1]。しかし、一般に食品の流通においては、加工の際に廃棄される部位があるため、加工後に流通量が減少する。従来方式では、この加工による流通量の減少の範囲内で、産地・銘柄が異なる食品を混入する偽装を検出できない。

また、このような偽装を検出するには、過去の傾向から逸脱する異常値を検出する手法が用いられる。例えば、正常時の値の変化を学習した予測モデルによる異常検知方式がある。食品トレーサビリティに応用し、加工後の重量の入力値と予測モデルの予測値の差(以下、予測誤差)に対して閾値を用いることで偽装を検出できる。しかし、産地・銘柄毎に加工後の重量のバラツキが異なることが想定されるため、産地・銘柄毎に閾値を作成する必要があり、産地・銘柄の組み合わせが大量に存在する食品においては、閾値作成の労力は許容できない。

3. 課題

閾値を用いずに加工による流通量の減少の範囲内で産地・銘柄が異なる食品を混入する偽装を検出することを課題とした。従来の異常検知方式の精度を目標値とした。

4. 提案システム

4.1 解決アイデア

食品は加工の際に廃棄される部位があるため、加工後に重量が減少する。我々は、食品の加工前後における重量の差が、産地・銘柄の組み合わせに対して固有の分布に従うものと仮定した。産地・銘柄の組み合わせ毎に食品の加工後の重量を予測し、予測誤差が最も小さい産地・銘柄の組み合わせを特定し、特定した産地・銘柄の組み合わせと生産地で登録された産地・銘柄の組み合わせが一致するかを確認することで、閾値を用いずに偽装を検出できると考えた。加工後の重量が偽装されていない場合は、生産地で登録された産地・銘柄の組み

合わせの予測誤差が他の産地・銘柄の組み合わせの予測誤差よりも小さくなり、正常と判定できる。加工後の重量が偽装された場合は、生産地で登録された産地・銘柄の組み合わせの予測誤差が大きくなり、他の産地・銘柄の組み合わせの予測誤差の方が小さくなり、異常と判定できる。

4.2 取扱履歴

取扱履歴とは、食品事業者間での食品の入出荷の履歴と事業者内での食品への生産・加工に関する履歴からなる。取扱履歴の作成時に入力する項目を生産、出荷、入荷、加工の工程別に策定した(表1)。

表1 工程別入力項目

工程	入力項目
生産	識別番号, 日時, 産地, 銘柄, 重量
出荷	識別番号, 日時, 出荷先, 重量
入荷	識別番号, 日時, 仕入れ元, 重量
加工	識別番号, 日時, 加工内容, 加工後の重量

4.3 システム構成

本システムは以下の2つの処理とからなる(図1)。

オフライン処理：過去の取扱履歴を取得し、産地、銘柄、加工前の重量、加工後の重量からなる学習データを生成する。学習データに含まれる加工後の重量を目的変数とし、銘柄、加工前の重量を説明変数とし、重回帰分析による産地毎の重回帰式を生成する。また、学習データに含まれる産地・銘柄の組み合わせを抽出する。

オンライン処理：食品事業者から入力を基に取扱履歴を生成する。オフライン処理で生成された重回帰式と産地・銘柄の組み合わせを用いて、全通りの産地・銘柄の組み合わせ毎の加工後の重量を予測する。食品事業者の入力値と最も予測値に近い産地・銘柄の組み合わせを特定する。特定した産地・銘柄の組み合わせと生産工程で登録された産地・銘柄の組み合わせが一致する場合に正常、一致しない場合に異常と判定し、判定結果を監査担当者へ通知する。正常と判定された取扱履歴は共有され、ブロックチェーン上で分散管理される。消費者は食品の識別番号を入力し、取扱履歴を追跡する。

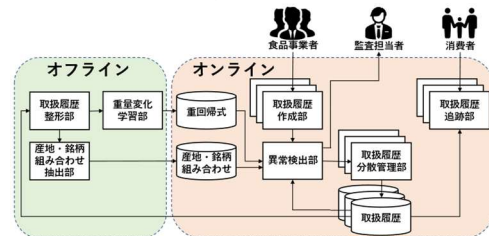


図1 システム構成図

5. 評価実験

5.1 目的と方法

提案方式の特性を検証するために、産地・銘柄の組み合わせ毎の固有の分布の差と加工後の重量に対する産地・銘柄が異なる食品を混入する量の割合(以下、混入割合)を変化させ、従来方式と提案方式で異常検出し、精度を比較した。

肉や魚、野菜などの食品の種類によって、食品の加工前後の重量差が産地と銘柄にどのように影響を受けるかは異なるこ

Development of a Traceability System Based on Consistency Determination of Transaction History to Detect Food Counterfeiting

[†]Shion Yamada, [†]Mirei Ishizuka, [†]Shoki Harada,
[†]Ichitoshi Takehara, [‡]Norihiko Kamata, [†]Koji Kida
[†]Kagawa University, [‡]NEC Communication Systems, Ltd.

とが想定される。本実験では加工前の重量の分布は産地に影響され、加工による廃棄割合の分布は銘柄に影響されると仮定し、擬似的な取扱履歴を生成した。

評価軸としては、誤検知よりも検知漏れがないことが重要となるが、正常な取扱履歴を異常と判定することが多いと、監査担当者が大量に確認しなければならなくなってしまうため、再現率と真陰性率の2つを設定した。

実験手順：産地毎の加工前の重量の分布の差と銘柄毎の廃棄割合の分布の差を変化させながら、繰り返した。

- ① 3つの産地に対応付けた3つの加工前の重量の分布と3つの銘柄に対応付けた3つの廃棄割合の分布を生成
- ② ①で生成した分布に従う乱数を生成し、加工前の重量と廃棄割合の積から加工後の重量を求めることで、産地、銘柄、加工前の重量、加工後の重量からなるサンプルデータを生成
- ③ ②で生成したサンプルデータの半分を学習データとして重回帰分析し、重回帰式を生成
- ④ 学習データと重回帰式から従来方式の閾値を設定
- ⑤ ②で生成したサンプルデータの残り半分を正常データとし、それを基に廃棄量を偽った異常データを生成
- ⑥ 正常データと異常データに対し、従来方式と提案方式で異常検出し、再現率と真陰性率を算出

サンプルデータ生成方法(①・②)：分布の差は、分布同士の重なりが少ないほど大きくなり、重なりが多いほど小さくなる。分布の重なりは平均値の差だけではなく、分布の分散によって変化するため、平均値の差と分散を加味した分布間の差の指標であるCohen's dを用いる。

加工前の重量分布と加工による廃棄割合の分布をそれぞれ隣り合う分布間のCohen's dが0から20へ0.5刻みで変化するように分散を固定し、平均値の差を変化させ生成する(図2)。生成した分布に従う乱数を基にサンプルデータを生成する。

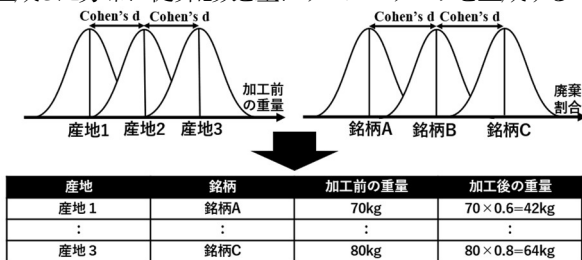


図2 サンプルデータ生成アルゴリズム

従来方式の閾値設定方法(④)：特定の確率分布を仮定せずに用いることができる汎用的な閾値設定の手法である分位点法を用いる。学習データに含まれる説明変数の組を重回帰式に入力し、予測した加工後の重量の値と学習データに含まれる加工後の重量の値から予測誤差を算出する。算出した予測誤差のうち、分位点95%となる値を閾値とする。

異常データの生成方法(⑤)：異常データは、正常データの混入割合を5%から30%へ5%刻みで変化させた6つのバリエーションで生成した。

5.2 実験結果

図3に加工前の重量の分布と廃棄割合の分布のCohen's dによる真陰性率、図4に混入割合による一方の方式よりも再現率を上回る条件の数を示す。

真陰性率が低い結果から、正常な取扱履歴を異常と判定する偽陽性の数が多いといえる。特に廃棄割合の分布間のCohen's dが0から5の区間では、曲面の傾きが急となっており、5から15の区間では緩やかになっていった。混入割合が5%から

15%の場合に再現率が高い結果から、混入割合が少ない場合に従来方式よりも偽装された取扱履歴を取りこぼし少なく異常と判定できているといえる。混入割合が20%から30%の場合に再現率が低い結果から偽装された取扱履歴を正常と判定する偽陰性が多く発生しているといえる。

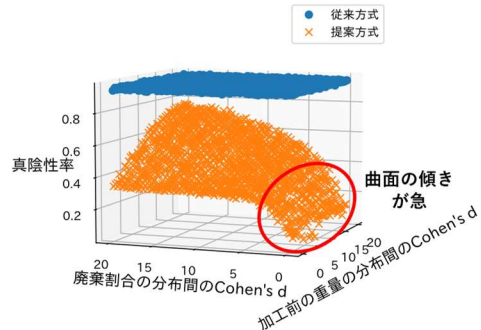


図3 加工前の重量の分布と廃棄割合の分布のCohen's dによる真陰性率

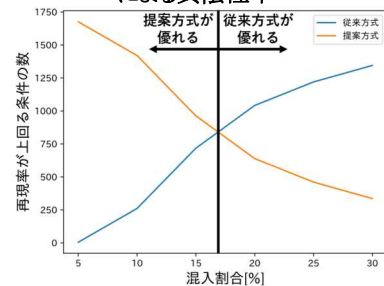


図4 混入割合による一方の方式よりも再現率を上回る条件の数

5.3 考察

再現率は混入割合が5%から15%の場合に目標値を達成したが、20%から30%の場合に未達成であった。真陰性率は目標値を未達成であった。その理由を考察する。

偽陰性の発生原因の一つとして、予測値が最も大きい産地・銘柄の食品に対し、産地・銘柄が異なる食品を混入された場合に常に予測値が最も近い産地・銘柄と生産地で入力された産地・銘柄が一致することが考えられる。そのため、産地・銘柄が異なる食品を混入しても異常と判定されない。最も予測値が大きい産地・銘柄の食品に対しては閾値を設定し、異常と判定できるようにすることで再現率を高められると考える。

偽陽性の発生原因の一つとして、異なる産地・銘柄の組み合わせであっても、対応する重回帰式が類似していることが考えられる。そのため、正常な取扱履歴であっても他の産地・銘柄の組み合わせの予測値の方が近くなってしまうため、異常と判定されてしまう。重回帰式が類似する産地・銘柄の組み合わせを同一グループとして扱い、同一グループの産地・銘柄の組み合わせが最も予測値が近くなった場合に正常と判定することにより、真陰性率を高められると考える。

6. おわりに

混入割合が5%から15%の場合に従来の異常検出方式よりも取りこぼし少なく偽装された取扱履歴を検出できることの結果を示した。また、廃棄割合の分布間のCohen's dが0から5の区間では正常な取扱履歴を異常と判定しやすいことを示した。

参考文献

- [1] 松村康弘, 加藤あすか, 三上貞芳, “食品トレーサビリティにおける流通フローの高精度化による信頼性向上に関する研究”, 情報処理学会第74回全国大会講演論文集, pp615-pp616, (2012)